A Remark on Estimating The Mean of A Normal Distribution with Known Coefficient of Variation

Rasul A. Khan
Cleveland State University, R.KHAN@csuohio.edu

A remark on estimating the mean of a normal distribution with known coefficient of variation

Rasul A. Khan

Let $X_1, X_2, \ldots, X_n$ be iid $N(\mu, a\mu^2)$ $(a > 0)$ random variables with an unknown mean $\mu > 0$ and known coefficient of variation (CV) $\sqrt{a}$. The estimation of μ is revisited and it is shown that a modified version of an unbiased estimator of μ [cf. Khan RA. A note on estimating the mean of a normal distribution with known CV. J Am Stat Assoc. 1968;63:1039–1041] is more efficient. A certain linear minimum mean square estimator of Gleser and Healy [Estimating the mean of a normal distribution with known CV. J Am Stat Assoc. 1976;71:977–981] is also modified and improved. These improved estimators are being compared with the maximum likelihood estimator under squared-error loss function. Based on asymptotic consideration, a large sample confidence interval is also mentioned.

Keywords: coefficient of variation; linear; maximum likelihood estimator; minimum squared-error; normal; unbiased; efficiency

Introduction

Let $X_1, X_2, \ldots, X_n$ be iid $N(\mu, \sigma^2)$ random variables with an unknown mean μ and unknown variance σ^2. There are several examples of practical problems in which the coefficient of variation (CV) σ/μ has a known constant value (see [1–3]). The problem of estimating μ was discussed by the author (cf. [4]) when the CV is known, and a certain linear unbiased (LU) estimator was explored. But there is a lacuna in the given estimator by Khan or linear minimum mean square (LMMS) estimator given by Gleser and Healy.[2] This will be clarified a bit more in the sequel. Therefore, the main reason for revisiting this problem is to find some more efficient estimators, and compare them with the maximum likelihood estimator (MLE).

Before pointing out the deficiencies in LU and LMMS estimators, and finding their improved competitors, we make the following observation which is the main reason for revisiting this problem. Since we are assuming that the CV (σ/μ) is known, this implies that the sign of μ is also known. Therefore, there is no loss of generality in assuming that $\mu > 0$ and $\sigma = \mu\sqrt{a}$, where a is a known positive constant. Thus $X_1, X_2, \ldots, X_n$ $(n \geq 2)$ are iid $N(\mu, a\mu^2)$ random variables, and it is of interest to find an efficient estimator of μ. Let

$$T_1^* = \bar{X} = \frac{X_1 + \cdots + X_n}{n}, \quad S^2 = n^{-1}\sum_1^n (X_i - \bar{X})^2, \quad T_2 = c_n S, \tag{1}$$

where $c_n = \sqrt{n}\Gamma((n-1)/2)/\sqrt{2a}\Gamma(n/2)$.

For some discussions in the sequel it is useful to define the squared-error loss function for the estimator $\hat{\mu}$ by

$$L(\hat{\mu}, \mu) = (\hat{\mu} - \mu)^2. \tag{2}$$

It is known [4] that $ET_1^* = ET_2 = \mu$. Moreover, it is also known that although $(\bar{X}, S)$ continues to be minimal sufficient, but it is not complete. Hence a unique UMVU (uniformly minimum variance unbiased) estimator of μ cannot be found (see [5]). However, still the problem remains to find a good and efficient estimator of μ.

Khan [4] considered a class of unbiased estimators linear in T_1^* and T_2 (to be designated as LU), and found the one with the minimum variance. That is, among all estimators of the form $\alpha T_1^* + (1-\alpha)T_2$, the best one was found, and was compared with the MLE for its asymptotic efficiency. It was shown that both estimators are best asymptotically normal (BAN). Gleser and Healy [2] considered the class C of estimators linear in T_1^* and T_2 of the form $c_1T_1^* + c_2T_2$ but not necessarily unbiased, and found the one with uniformly minimum mean squared-error (LMMS). However, despite the fact that these have minimum risk and are BAN estimators, both (LU and LMMS) have the flaw of being possibly negative with positive probability for estimating a positive parameter. In fact, both are inadmissible under squared-error loss (2) (inadmissibility of LMMS was pointed out by Gleser and Healy [2]) as is easily seen by comparing the risks of these estimators with their positive parts. The use of the positive part does not really remedy the situation. In case the positive part is zero, it cannot be used as an estimate of $\mu > 0$.

Gleser and Healy [2] have noted that μ is a scale parameter, and all the estimators considered (LU, LMMS and MLE) are scale equivariant. Motivated by this observation, they obtained the minimum risk scale equivariant estimator. However, the estimator may be difficult to use due to its inherent complexity. Besides the difficulty of using the estimator, it is not even known how much reduction in risk is achieved in comparison to LU, LMMS or MLE. In fact, the associated risk of the scale equivariant estimator is completely intractable, and the risk reduction in its use cannot be determined. The Bayes estimator has similar drawback, and no favourable argument can be made when its use has the same computational difficulty. Therefore, the best competing estimators are LU, LMMS, and the MLE because of their known properties and the evaluation of exact risks. Consequently, the purpose of this note is to modify the LU and LMMS estimators to make them more efficient, and make numerical comparisons of their risks with the MLE. Based on the modified LU estimator, a large sample confidence interval for μ is also described.

The main result

Recall from Equation (1) the usual unbiased estimators $\bar{X}$ and c_nS of μ as

$$T_1^* = \bar{X}, \quad T_2 = c_nS, \quad c_n = \frac{\sqrt{n}\Gamma((n-1)/2)}{\sqrt{2a}\Gamma(n/2)}.$$

Also, define

$$V_1^* = \frac{a}{n}, \quad V_2 = \frac{a(n-1)}{n}c_n^2 - 1 = \frac{(n-1)}{2}\frac{\Gamma^2((n-1)/2)}{\Gamma^2(n/2)} - 1. \tag{3}$$

It is known [4] that

$$ET_1^* = ET_2 = \mu, \quad \sigma^2_{T_1^*} = \mu^2 V_1^*, \quad \sigma^2_{T_2} = \mu^2 V_2.$$

Using the definitions in Equation (3), let

$$d = \alpha T_1^* + (1-\alpha)T_2, \quad \alpha = \frac{V_2}{V_1^* + V_2}. \tag{4}$$

It has been shown in [4] that d is the best LU estimator of μ, and its minimum variance is

$$\sigma_d^2 = \frac{\mu^2 V_1^* V_2}{V_1^* + V_2}. \tag{5}$$

Now we will modify this and the LMMS estimator to improve their efficiencies. Let $T_1 = \beta|\bar{X}|$ with a suitable $\beta > 0$. It is easy to verify that

$$E|\bar{X}| = \mu\left(2\Phi\left(\sqrt{\frac{n}{a}}\right) - 1 + \sqrt{\frac{2a}{n\pi}}\exp\left(-\frac{n}{2a}\right)\right),$$

where $\Phi(\cdot)$ is the standard normal distribution function. Let the normal density function be $\phi(x) = \exp(-x^2/2)/\sqrt{2\pi}$ and $\lambda = \lambda_n = \sqrt{n/a}$ (for notational simplicity we may suppress the dependence of λ or β on n), and define

$$\beta = \beta_n = \beta(\lambda_n) = \frac{1}{2\Phi(\lambda) - 1 + (2/\lambda)\phi(\lambda)}. \tag{6}$$

This choice of β gives $ET_1 = \mu$. Moreover, it is easy to see that

$$\sigma_{T_1}^2 = \mu^2 V_1, \quad \text{where } V_1 = (\beta_n^2 - 1) + \frac{a\beta_n^2}{n}. \tag{7}$$

For later use, we observe the asymptotic behaviour of β_n and V_1 as $n \to \infty$. Recall the definition of β_n by Equation (6) where $\lambda_n = \sqrt{n/a}$. It is obvious from Equation (6) that $\beta_n \to 1$ as $n \to \infty$. However, using the fact that $n(1 - \Phi(\sqrt{n/a})) \to 0$ as $n \to \infty$, it can be verified that $n(\beta_n^2 - 1) \to 0$ as $n \to \infty$. Consequently, it follows from Equation (7) that $nV_1 \to a$ as $n \to \infty$.

Now consider the class of unbiased estimators defined by

$$d^* = cT_1 + (1-c)T_2, \quad 0 \le c \le 1.$$

Since $\sigma_{T_2}^2 = \mu^2 V_2$, using Equations (3) and (7) we have

$$\sigma_{d^*}^2 = \mu^2(c^2 V_1 + (1-c)^2 V_2),$$

and it is minimized by choosing $c = V_2/(V_1 + V_2)$, and the minimum variance is

$$\sigma_{d^*}^2 = \frac{\mu^2 V_1 V_2}{V_1 + V_2}. \tag{8}$$

Thus d^* is the modified best unbiased estimator of μ which is linear in $|\bar{X}|$ and S. We will now show that d^* uniformly dominates d defined by Equation (4). To this end, we first show that $\beta(\lambda) < 1$ for every $\lambda > 0$ and all n. Let

$$F(\lambda) = 2\Phi(\lambda) - 1 + \frac{2}{\lambda}\phi(\lambda), \quad \text{then } \beta(\lambda) = \frac{1}{F(\lambda)}.$$

Clearly,

$$F'(\lambda) = -\frac{2}{\lambda^2}\phi(\lambda) \quad \text{and} \quad \beta'(\lambda) = \frac{2\phi(\lambda)}{(\lambda F(\lambda))^2} > 0.$$

Thus $\beta(\lambda)$ is increasing in $\lambda > 0$, and since $\lim_{\lambda\to\infty}\beta(\lambda) = 1$, hence $\beta(\lambda) < 1$ for every $\lambda > 0$. Consequently, $V_1 < V_1^*$, where V_1^* and V_1 are defined by Equations (3) and (7). Now going back to Equation (8) we see that

$$\sigma_{d^*}^2 = \mu^2 g(V_1), \quad g(V_1) = \frac{V_1 V_2}{V_1 + V_2}.$$

Since $g'(V_1) = (V_2/(V_1+V_2))^2 > 0$, $g(V_1)$ is increasing in V_1, and we conclude that $g(V_1) < g(V_1^*)$. Thus, it follows from Equations (5) and (8) that $\sigma_{d^*}^2 < \sigma_d^2$ uniformly in μ. It should also be emphasized that d^* continues to be BAN for the same reasons as given in [4] combined with the earlier noted fact that $n\sigma_{T_1}^2 \to a\mu^2$ as $n \to \infty$.

Now we will provide a modified version of the minimum mean squared-error estimator of Gleser and Healy.[2] Let T_1 and T_2 be as before and consider the class of estimators of the form $c_1T_1 + c_2T_2$ (i.e. linear in T_1 and T_2 but not necessarily unbiased). It follows from Lemma 2.1 in [2, p.978] that the best linear estimator being considered here is given by

$$T = c_1^* T_1 + c_2^* T_2, \quad \text{where } c_1^* = \frac{V_2}{V_1 + V_2 + V_1V_2}, \quad c_2^* = \frac{V_1}{V_1 + V_2 + V_1V_2}.$$

The associated minimum risk is given by

$$R(T,\mu) = \mu^2 g_1(V_1) = \mu^2 \left(\frac{V_1V_2}{V_1 + V_2 + V_1V_2} \right). \tag{9}$$

Since $g_1'(V_1) = (V_2/(V_1 + V_2 + V_1V_2))^2 > 0$, hence $g_1(V_1)$ is increasing and $g_1(V_1) < g_1(V_1^*)$, where V_1^* has been defined in Equation (3). Thus, it follows that T is uniformly better than $\hat{\mu}_{\text{LMMS}}$ of Gleser and Healy [2, see Equation (2.6), p.978].

Remark The estimators d^* and T are scale equivariant, and like d^*, T is also BAN estimator.

Now we will compare the modified best unbiased estimator d^*, the modified best LMMS T and the MLE estimator in terms of their risks under squared-error loss for some values of (a, n). For the sake of completeness, we will describe the associated risk of the MLE. To this end, we note that the MLE as given in [4, p.1041] or [2, p.978] is

$$\hat{\mu} = \frac{-\bar{X} + \sqrt{4aS^2 + (1+4a)\bar{X}^2}}{2a}.$$

To describe the associated risk let $m = n - 1$ and define

$$f(u,v) = \frac{1}{\sqrt{2\pi}2^{m/2}\Gamma(m/2)} v^{m/2-1} \exp\left(-\frac{u^2+v}{2}\right).$$

It should be noted that $f(u,v)$ is the product of the standard normal density and the gamma density functions. Clearly, the associated risk is $R(\hat{\mu}) = R(\hat{\mu},\mu) = E(\hat{\mu}-\mu)^2$. Now write the expectation with respect to the distributions of $\bar{X}$ as normal $N(\mu, a\mu^2/n)$ and $nS^2/a\mu^2$ as $\chi^2(n-1)$, it can be simplified as follows. Let $\sigma^2 = a\mu^2$, $v = (nS^2/\sigma^2)$ and $K = \sqrt{n/2\pi\sigma^2}(1/2^{m/2}\Gamma(m/2))$.

Table 1. Risks of d^*, T and $\hat{\mu}$.

	$a=1$			$a=2$		
n	$R_1(d^*)$	$R_2(T)$	$R_3(\hat{\mu})$	$R_1(d^*)$	$R_2(T)$	$R_3(\hat{\mu})$
4	0.1001	0.0910	0.0843	0.1191	0.1064	0.0994
9	0.0406	0.0391	0.0373	0.0492	0.0469	0.0444
16	0.0219	0.0215	0.0209	0.0266	0.0259	0.0250
25	0.0138	0.0136	0.0134	0.0167	0.0164	0.0160

Also, let

$$Q(x,v) = \left(\frac{-x + \sqrt{(4a\sigma^2 v/n) + (1+4a)x^2}}{2a} - \mu \right)^2 .$$

Then, it follows that

$$R(\hat{\mu}) = K \int_0^\infty \int_{-\infty}^\infty Q(x,v) v^{m/2-1} \exp\left(-\frac{n}{2\sigma^2}(x-\mu)^2 + \frac{v}{2} \right) \mathrm{d}x\, \mathrm{d}v.$$

Making the substitution $u = \sqrt{n/\sigma^2}(x-\mu)$ one verifies that

$$R(\hat{\mu}) = \frac{\mu^2}{4a^2} \int_0^\infty \int_{-\infty}^\infty g(u,v) f(u,v)\, \mathrm{d}u\, \mathrm{d}v, \tag{10}$$

where $f(u,v)$ has been defined above and

$$g(u,v) = \left(\sqrt{\left(\frac{4a^2}{n}\right) v + (1+4a)\left(1 + u\sqrt{\left(\frac{a}{n}\right)}\right)} - \left(1 + u\sqrt{\left(\frac{a}{n}\right)}\right) - 2a \right)^2 .$$

Unfortunately, $R(\hat{\mu})$ in Equation (10) cannot be evaluated explicitly. Therefore, its numerical computation uses Maple software. Now we will compare these competing estimators numerically in terms of their risks for certain values of (n,a). However, since μ^2 is a multiplying factor in all three risks, we simply compare the risks divided by μ^2. Thus we use the following abbreviated notations. Let $R_1(d_1^*) = \mu^{-2}\sigma^2_{d^*}$, $R_2(T) = \mu^{-2}R(T,\mu)$ and $R_3(\hat{\mu}) = \mu^{-2}R(\hat{\mu})$ (defined by (10)) be the risks of the modified best LU, the modified best LMMS and the MLE, respectively. Table 1 gives the values of these risks for certain values of n and $a = 1, 2$.

Conclusions

The calculations show that the differences in the three risks are very minor. All the risks are practically the same for almost all n. All three estimators enjoy the same asymptotic properties although d^* has the advantage of being unbiased. As far as scale equivariant estimator or Bayes estimator in [2, p.979] is concerned, both have the drawback of computational difficulty. These cannot be implemented easily even with the aid of a computer. Moreover, it is practically impossible to find the associated risk, and the ensuing reduction cannot be assessed. The suspicion is that neither will have any significant reduction in risk. An infinitesimal risk reduction will certainly undermine the complex task of using any of these estimators. As far as the modified LU, LMMS and the MLE are concerned, one can use any of these depending on one's preference. Thus d^*, T or the MLE are all the best alternatives.

We conclude by a brief discussion of a confidence interval for μ. It can be easily verified that the distribution of d^*/μ is independent of μ. However, an exact distribution of d^*/μ in a closed and numerically useable form is not possible for obvious reasons. The best one can do is an infinite series distribution involving incomplete gamma functions. Therefore, we only consider a large sample approximate confidence interval based on d^*. In what follows we let $\Phi(k) = 1 - \frac{1}{2}\gamma$, $0 < \gamma < 1$. Since d^* is BAN, we have

$$\frac{\sqrt{n(1+2a)}(d^* - \mu)}{\mu\sqrt{a}} \xrightarrow{\mathcal{D}} N(0,1) \quad \text{as } n \to \infty.$$

Hence, a large sample approximate $(1-\gamma)100\%$ confidence interval for μ has its lower and upper bounds as

$$L = \frac{d^*}{1 + k\sqrt{a/n(1+2a)}}, \quad U = \frac{d^*}{1 - k\sqrt{a/n(1+2a)}}.$$

Finally, we close the discussion by a remark that one can use either LMMS (T) or the MLE in the above bounds as these estimators have exactly the same asymptotic distribution.

References

[1] Amemiya T. Regression analysis when the variance of the dependent variable is proportional to the square of its expectation. J Am Stat Assoc. 1973;68:928–934.

[2] Gleser LJ, Healy JD. Estimating the mean of a normal distribution with known coefficient of variation. J Am Stat Assoc. 1976;71:977–981.

[3] Khan RA. A note on testing the means of normal distributions with known coefficient of variations. Commun Stat Theory Methods A. 1978;7(9):867–876.

[4] Khan RA. A note on estimating the mean of a normal distribution with known coefficient of variation. J Am Stat Assoc. 1968;63:1039–1041.

[5] Lehmann EL. Theory of point estimation. Pacific Grove (CA): Wadsworth; 1991.